

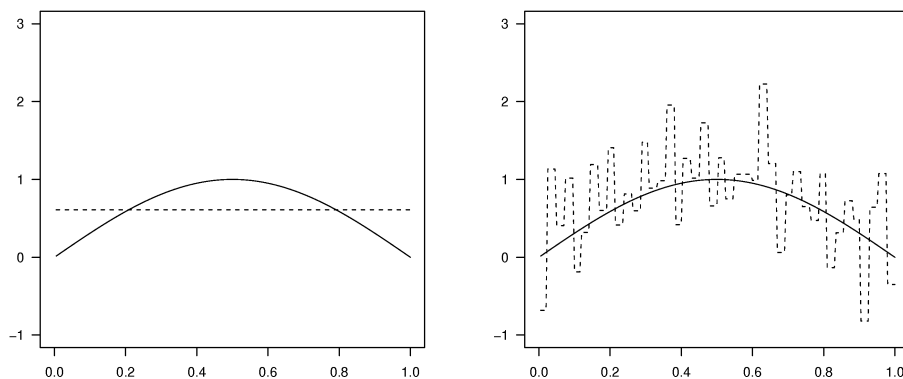
A course on asymptotic methods, choice of model in regression and causality

Jean-Marc Azaïs, University of Toulouse.

This course is an introduction to methods of choice of variables in multiple regression models under non-normal assumptions.

After a reminder on linear models and in particular on orthogonality, we present asymptotic properties of estimators and tests for large dimensions linear models under non-normality.

The over-fitting and the necessity of choice of variables is illustrated in the following figure where a piecewise constant estimation over 42 sub-intervals is used :



The course presents criterions of choice of models : AIC, BIC, CP, PRESS, and their relation with tests and thresholding. Convergence of the choice by AIC, BIC, CP under non Gaussian assumption is established

We will finish by a short presentation of sparse method as LASSO and make a review on causality including randomized experiments, Austin Bradford Hill definition of causality, Granger causality and intervention calculus : IDA.

References

J-M Azaïs asymptotic selection of models. Preprint

J-M Azaïs & J-M Bardet le modèle linéaire par l'exemple. Dunod (2nd edition) (in french)

Hill, A. B. (1965). "The Environment and Disease: Association or Causation?". Proceedings of the Royal Society of Medicine 58 (5): 295-300.

Granger, C.W.J., 1969. "Investigating causal relations by econometric models and cross-spectral methods". Econometrica 37 (3), 424-438.

Bühlmann, P. (2011). Causal statistical inference in high dimensions. Preprint.